

1009

#2 OIPE

RAW SEQUENCE LISTING DATE: 10/17/2001
PATENT APPLICATION: US/09/842,347 TIME: 14:55:22

Input Set : N:\Crf3\RULE60\09842347.txt
Output Set: N:\CRF3\10172001\I842347.raw

SEQUENCE LISTING

ENTERED

4 (1) GENERAL INFORMATION:

6 (i) APPLICANT: TAKAHASHI, Tohru
7 SERIZAWA, Nobufusa
8 KOISHI, Ryuta
9 KAWASHIMA, Ichiro

11 (ii) TITLE OF INVENTION: EXPRESSION SYSTEMS UTILIZING
12 AUTOLYZING FUSION PROTEINS
13 AND A NOVEL REDUCING POLYPEPTIDE

15 (iii) NUMBER OF SEQUENCES: 19

17 (iv) CORRESPONDENCE ADDRESS:

18 (A) ADDRESSEE: Frishhauf, Holtz, Goodman, Langer & Chick, P.C.
19 (B) STREET: 767 Third Avenue-25th Floor
20 (C) CITY: New York
21 (D) STATE: New York
22 (E) COUNTRY: United States
23 (F) ZIP: 10017-2023

25 (v) COMPUTER READABLE FORM:

26 (A) MEDIUM TYPE: Floppy disk
27 (B) COMPUTER: IBM PC compatible
28 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
29 (D) SOFTWARE: PatentIn Release #1.24

31 (vi) CURRENT APPLICATION DATA:

C--> 32 (A) APPLICATION NUMBER: US/09/842,347
C--> 33 (B) FILING DATE: 25-Apr-2001
34 (C) CLASSIFICATION:

44 (vii) PRIOR APPLICATION DATA:

37 (A) APPLICATION NUMBER: 09/167,151
38 (B) FILING DATE: 1998-10-06
41 (A) APPLICATION NUMBER: JP 6-218392
42 (B) FILING DATE: 13-SEP-1994
45 (A) APPLICATION NUMBER: JP 6-303809
46 (B) FILING DATE: 07-DEC-1994

48 (viii) ATTORNEY/AGENT INFORMATION:

49 (A) NAME: Goodman, Herbert
50 (B) REGISTRATION NUMBER: 17081
51 (C) REFERENCE/DOCKET NUMBER: 950376/HG

53 (ix) TELECOMMUNICATION INFORMATION:

54 (A) TELEPHONE: (212) 319-4900
55 (B) TELEFAX: (212) 319-5101
56 (C) TELEX: 236268

59 (2) INFORMATION FOR SEQ ID NO: 1:

61 (i) SEQUENCE CHARACTERISTICS:

62 (A) LENGTH: 1320 base pairs
63 (B) TYPE: nucleic acid
64 (C) STRANDEDNESS: double
65 (D) TOPOLOGY: linear

RAW SEQUENCE LISTING

DATE: 10/17/2001

PATENT APPLICATION: US/09/842,347

TIME: 14:55:22

Input Set : N:\Crf3\RULE60\09842347.txt

Output Set: N:\CRF3\10172001\I842347.raw

```

67      (ii) MOLECULE TYPE: cDNA to mRNA
W--> 69      (iii) HYPOTHETICAL: N
W--> 71      (iv) ANTI-SENSE: N
73      (vi) ORIGINAL SOURCE:
74          (A) ORGANISM: Clover Yellow Vein Virus
76      (ix) FEATURE:
77          (A) NAME/KEY: CDS
78          (B) LOCATION: 1..1320
79          (D) OTHER INFORMATION:
80      (ix) FEATURE:
81          (A) NAME/KEY: mat_peptide
82          (B) LOCATION: 10..1311
83          (D) OTHER INFORMATION:
87      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
89 AAG TTC CAA GGG AAA AGT AAG AGA ACA AGA CAA AAG TTG AAG TTC AGA      48
90 Lys Phe Gln Gly Lys Ser Lys Arg Thr Arg Gln Lys Leu Lys Phe Arg
91 1      5      10      15
93 GCG GCA AGA GAC ATG AAG GAT CGT TAT GAA GTG CAT GCC GAT GAG GGG      96
94 Ala Ala Arg Asp Met Lys Asp Arg Tyr Glu Val His Ala Asp Glu Gly
95      20      25      30
97 ACT TTA GTG GAA AAT TTT GGA ACT CGT TAT TCA AAG AAA GGC AAG ACA      144
98 Thr Leu Val Glu Asn Phe Gly Thr Arg Tyr Ser Lys Lys Gly Lys Thr
99      35      40      45
101 AAA GGT ACT GTT GTG GGT TTG GGT GCA AAA ACA AGA CGG TTC ACT AAC      192
102 Lys Gly Thr Val Val Gly Leu Gly Ala Lys Thr Arg Arg Phe Thr Asn
103      50      55      60
105 ATG TAT GGT TTT GAC CCC ACG GAG TAT TCA TTT GCT AGG TAT CTT GAT      240
106 Met Tyr Gly Phe Asp Pro Thr Glu Tyr Ser Phe Ala Arg Tyr Leu Asp
107 65      70      75      80
109 CCA ATC ACG GGT GCA ACA TTG GAT GAA ACC CCA ATT CAC AAT GTA AAT      288
110 Pro Ile Thr Gly Ala Thr Leu Asp Glu Thr Pro Ile His Asn Val Asn
111      85      90      95
113 TTG GTT GCT GAG CAT TTT GGC GAC ATA AGG CTT GAT ATG GTT GAC AAG      336
114 Leu Val Ala Glu His Phe Gly Asp Ile Arg Leu Asp Met Val Asp Lys
W--> 115      100      105      110
117 GAG TTA CTT GAC AAA CAG CAC TTA TAC CTC AAG AGA CCA ATA GAA TGT      384
118 Glu Leu Leu Asp Lys Gln His Leu Tyr Leu Lys Arg Pro Ile Glu Cys
W--> 119      115      120      125
121 TAC TTT GTA AAG GAT GCT GGT CAG AAG GTG ATG AGG ATT GAT CTA ACA      432
122 Tyr Phe Val Lys Asp Ala Gly Gln Lys Val Met Arg Ile Asp Leu Thr
W--> 123      130      135      140
125 CCC CAC AAC CCA TTG TTG GCA AGC GAT GTT AGC ACA ACC ATA ATG GGT      480
126 Pro His Asn Pro Leu Leu Ala Ser Asp Val Ser Thr Thr Ile Met Gly
W--> 127 145      150      155      160
129 TAT CCT GAG AGA GAA GGT GAA CTC CGT CAA ACT GGA AAG GCA AGG TTA      528
130 Tyr Pro Glu Arg Glu Gly Glu Leu Arg Gln Thr Gly Lys Ala Arg Leu
W--> 131      165      170      175
133 GTC GAC CCA TCA GAG TTG CCC GCG CGG AAT GAG GAT ATT GAT GCA GAG      576
134 Val Asp Pro Ser Glu Leu Pro Ala Arg Asn Glu Asp Ile Asp Ala Glu

```

RAW SEQUENCE LISTING

DATE: 10/17/2001

PATENT APPLICATION: US/09/842,347

TIME: 14:55:22

Input Set : N:\Crf3\RULE60\09842347.txt

Output Set: N:\CRF3\10172001\I842347.raw

```

W--> 135          180          185          190
      137 TTT GAG AGT CTA AAT CGC ATA AGT GGT TTG CGC GAC TAT AAT CCC ATT      624
      138 Phe Glu Ser Leu Asn Arg Ile Ser Gly Leu Arg Asp Tyr Asn Pro Ile
W--> 139          195          200          205
      141 TCA CAA AAT GTT TGC TTG CTA ACA AAT GAG TCA GAA GGC CAT AGA GAG      672
      142 Ser Gln Asn Val Cys Leu Leu Thr Asn Glu Ser Glu Gly His Arg Glu
W--> 143          210          215          220
      145 AAG ATG TTT GGA ATT GGA TAT GGT TCA GTG ATC ATT ACA AAT CAA CAT      720
      146 Lys Met Phe Gly Ile Gly Tyr Gly Ser Val Ile Ile Thr Asn Gln His
W--> 147 225          230          235          240
      149 CTG TTC AGA AGG AAT AAT GGG GAG TTA TCA ATT CAA TCC AAG CAT GGC      768
      150 Leu Phe Arg Arg Asn Asn Gly Glu Leu Ser Ile Gln Ser Lys His Gly
W--> 151          245          250          255
      153 TAC TTC AGA TGC CGC AAC ACC ACA AGC TTG AAG ATG CTG CCT TTG GAG      816
      154 Tyr Phe Arg Cys Arg Asn Thr Thr Ser Leu Lys Met Leu Pro Leu Glu
W--> 155          260          265          270
      157 GGA CAT GAC ATT TTG TTG ATT CAG TTA CCA AGG GAC TTT CCA GTG TTT      864
      158 Gly His Asp Ile Leu Leu Ile Gln Leu Pro Arg Asp Phe Pro Val Phe
W--> 159          275          280          285
      161 CCA CAA AAG ATT CGC TTT AGG GAG CCA AGA GTG GAT GAC AAA ATT GTT      912
      162 Pro Gln Lys Ile Arg Phe Arg Glu Pro Arg Val Asp Asp Lys Ile Val
W--> 163          290          295          300
      165 TTG GTC AGC ACA AAT TTC CAG GAA AAG AGT TCC TCG AGC ACG GTC TCA      960
      166 Leu Val Ser Thr Asn Phe Gln Glu Lys Ser Ser Ser Thr Val Ser
W--> 167 305          310          315          320
      169 GAG TCC AGT AAC ATT TCA AGA GTG CAG TCA GCC AAT TTC TAC AAG CAT      1008
      170 Glu Ser Ser Asn Ile Ser Arg Val Gln Ser Ala Asn Phe Tyr Lys His
W--> 171          325          330          335
      173 TGG ATC TCA ACA GTA GCA GGA CAC TGT GGA AAC CCT ATG GTT TCG ACT      1056
      174 Trp Ile Ser Thr Val Ala Gly His Cys Gly Asn Pro Met Val Ser Thr
W--> 175          340          345          350
      177 AAA GAT GGA TTT ATT GTA GGT ATC CAC AGT CTT GCT TCA TTG ACA GGC      1104
      178 Lys Asp Gly Phe Ile Val Gly Ile His Ser Leu Ala Ser Leu Thr Gly
W--> 179          355          360          365
      181 GAC GTT AAC ATC TTC ACA AGC TTT CCG CCG CAG TTT GAG AAC AAA TAT      1152
      182 Asp Val Asn Ile Phe Thr Ser Phe Pro Pro Gln Phe Glu Asn Lys Tyr
W--> 183          370          375          380
      185 CTA CAG AAG CTC AGT GAA CAC ACA TGG TGT AGT GGA TGG AAA CTA AAT      1200
      186 Leu Gln Lys Leu Ser Glu His Thr Trp Cys Ser Gly Trp Lys Leu Asn
W--> 187 385          390          395          400
      189 CTT GGA AAG ATT AGT TGG GGT GGA ATC AAC ATT GTG GAG GAT GCA CCT      1248
      190 Leu Gly Lys Ile Ser Trp Gly Gly Ile Asn Ile Val Glu Asp Ala Pro
W--> 191          405          410          415
      193 GAA GAG CCC TTT ATA ACA TCC AAG ATG GCA AGC CTT CTT AGT GAT TTG      1296
      194 Glu Glu Pro Phe Ile Thr Ser Lys Met Ala Ser Leu Leu Ser Asp Leu
W--> 195          420          425          430
      197 AAT TGT TCA TTC CAA GCA AGT GCG      1320
      198 Asn Cys Ser Phe Gln Ala Ser Ala
W--> 199          435          440

```

RAW SEQUENCE LISTING

DATE: 10/17/2001

PATENT APPLICATION: US/09/842,347

TIME: 14:55:22

Input Set : N:\Crf3\RULE60\09842347.txt

Output Set: N:\CRF3\10172001\I842347.raw

202 (2) INFORMATION FOR SEQ ID NO: 2:

204 (i) SEQUENCE CHARACTERISTICS:

205 (A) LENGTH: 440 amino acids

206 (B) TYPE: amino acid

207 (D) TOPOLOGY: linear

209 (ii) MOLECULE TYPE: protein

211 (vi) ORIGINAL SOURCE:

212 (A) ORGANISM: Clover Yellow Vein Virus

214 (ix) FEATURE:

215 (A) NAME/KEY: mat_peptide

216 (B) LOCATION: 4..437

217 (D) OTHER INFORMATION:

220 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

222 Lys Phe Gln Gly Lys Ser Lys Arg Thr Arg Gln Lys Leu Lys Phe Arg
223   1           5           10           15
225 Ala Ala Arg Asp Met Lys Asp Arg Tyr Glu Val His Ala Asp Glu Gly
226           20           25           30
228 Thr Leu Val Glu Asn Phe Gly Thr Arg Tyr Ser Lys Lys Gly Lys Thr
229           35           40           45
231 Lys Gly Thr Val Val Gly Leu Gly Ala Lys Thr Arg Arg Phe Thr Asn
232           50           55           60
234 Met Tyr Gly Phe Asp Pro Thr Glu Tyr Ser Phe Ala Arg Tyr Leu Asp
235           65           70           75           80
237 Pro Ile Thr Gly Ala Thr Leu Asp Glu Thr Pro Ile His Asn Val Asn
238           85           90           95
240 Leu Val Ala Glu His Phe Gly Asp Ile Arg Leu Asp Met Val Asp Lys
241           100          105          110
243 Glu Leu Leu Asp Lys Gln His Leu Tyr Leu Lys Arg Pro Ile Glu Cys
244           115          120          125
246 Tyr Phe Val Lys Asp Ala Gly Gln Lys Val Met Arg Ile Asp Leu Thr
247           130          135          140
249 Pro His Asn Pro Leu Leu Ala Ser Asp Val Ser Thr Thr Ile Met Gly
250           145          150          155          160
252 Tyr Pro Glu Arg Glu Gly Glu Leu Arg Gln Thr Gly Lys Ala Arg Leu
253           165          170          175
255 Val Asp Pro Ser Glu Leu Pro Ala Arg Asn Glu Asp Ile Asp Ala Glu
256           180          185          190
258 Phe Glu Ser Leu Asn Arg Ile Ser Gly Leu Arg Asp Tyr Asn Pro Ile
259           195          200          205
261 Ser Gln Asn Val Cys Leu Leu Thr Asn Glu Ser Glu Gly His Arg Glu
262           210          215          220
264 Lys Met Phe Gly Ile Gly Tyr Gly Ser Val Ile Ile Thr Asn Gln His
265           225          230          235          240
267 Leu Phe Arg Arg Asn Asn Gly Glu Leu Ser Ile Gln Ser Lys His Gly
268           245          250          255
270 Tyr Phe Arg Cys Arg Asn Thr Thr Ser Leu Lys Met Leu Pro Leu Glu
271           260          265          270
273 Gly His Asp Ile Leu Leu Ile Gln Leu Pro Arg Asp Phe Pro Val Phe
274           275          280          285

```

RAW SEQUENCE LISTING

DATE: 10/17/2001

PATENT APPLICATION: US/09/842,347

TIME: 14:55:22

Input Set : N:\Crf3\RULE60\09842347.txt

Output Set: N:\CRF3\10172001\I842347.raw

```

276 Pro Gln Lys Ile Arg Phe Arg Glu Pro Arg Val Asp Asp Lys Ile Val
277      290      295      300
279 Leu Val Ser Thr Asn Phe Gln Glu Lys Ser Ser Ser Thr Val Ser
280 305      310      315      320
282 Glu Ser Ser Asn Ile Ser Arg Val Gln Ser Ala Asn Phe Tyr Lys His
283      325      330      335
285 Trp Ile Ser Thr Val Ala Gly His Cys Gly Asn Pro Met Val Ser Thr
286      340      345      350
288 Lys Asp Gly Phe Ile Val Gly Ile His Ser Leu Ala Ser Leu Thr Gly
289      355      360      365
291 Asp Val Asn Ile Phe Thr Ser Phe Pro Pro Gln Phe Glu Asn Lys Tyr
292      370      375      380
294 Leu Gln Lys Leu Ser Glu His Thr Trp Cys Ser Gly Trp Lys Leu Asn
295 385      390      395      400
297 Leu Gly Lys Ile Ser Trp Gly Gly Ile Asn Ile Val Glu Asp Ala Pro
298      405      410      415
300 Glu Glu Pro Phe Ile Thr Ser Lys Met Ala Ser Leu Leu Ser Asp Leu
301      420      425      430
303 Asn Cys Ser Phe Gln Ala Ser Ala
304      435      440

```

306 (2) INFORMATION FOR SEQ ID NO: 3:

308 (i) SEQUENCE CHARACTERISTICS:

309 (A) LENGTH: 25 base pairs

310 (B) TYPE: nucleic acid

311 (C) STRANDEDNESS: single

312 (D) TOPOLOGY: linear

314 (ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

W--> 316 (iii) HYPOTHETICAL: N

W--> 318 (iv) ANTI-SENSE: N

322 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

324 GTCCATGGGG AAAAGTAAGA GAACA

25

326 (2) INFORMATION FOR SEQ ID NO: 4:

328 (i) SEQUENCE CHARACTERISTICS:

329 (A) LENGTH: 20 base pairs

330 (B) TYPE: nucleic acid

331 (C) STRANDEDNESS: single

332 (D) TOPOLOGY: linear

334 (ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

W--> 336 (iii) HYPOTHETICAL: N

W--> 338 (iv) ANTI-SENSE: N

342 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

344 ACTCTGAGAC CGTGCTCGAG

20

346 (2) INFORMATION FOR SEQ ID NO: 5:

348 (i) SEQUENCE CHARACTERISTICS:

349 (A) LENGTH: 20 base pairs

350 (B) TYPE: nucleic acid

351 (C) STRANDEDNESS: single

352 (D) TOPOLOGY: linear

354 (ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

VERIFICATION SUMMARY

DATE: 10/17/2001

PATENT APPLICATION: US/09/842,347

TIME: 14:55:23

Input Set : N:\Crf3\RULE60\09842347.txt

Output Set: N:\CRF3\10172001\I842347.raw

L:32 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:33 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:69 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=1
L:71 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=1
L:115 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:119 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:123 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:127 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:131 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:135 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:139 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:143 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:147 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:151 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:155 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:159 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:163 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:167 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:171 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:175 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:179 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:183 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:187 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:191 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:195 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:199 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:316 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=3
L:318 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=3
L:336 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=4
L:338 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=4
L:356 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=5
L:358 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=5
L:376 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=6
L:378 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=6
L:396 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=7
L:398 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=7
L:416 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=8
L:418 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=8
L:436 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=9
L:455 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=10
L:474 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=11
L:476 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=11
L:769 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=13
L:771 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=13
L:789 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=14
L:791 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=14
L:811 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=15
L:813 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=15

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/842,347

DATE: 10/17/2001

TIME: 14:55:23

Input Set : N:\Crf3\RULE60\09842347.txt

Output Set: N:\CRF3\10172001\I842347.raw

L:831 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=16
L:833 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=16
L:851 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=17
L:853 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=17
L:871 M:246 W: Invalid value of Alpha Sequence Header Field, [HYPOTHETICAL:], SeqNo=18
L:873 M:246 W: Invalid value of Alpha Sequence Header Field, [ANTI-SENSE:], SeqNo=18
L:894 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:19